

# AFLOW: an automatic framework for high-throughput materials discovery

Stefano Curtarolo<sup>a,b,\*</sup>, Wahyu Setyawan<sup>a</sup>, Gus L. W. Hart<sup>c</sup>, Michal Jahnatek<sup>a</sup>, Roman V. Chepulskii<sup>a</sup>, Richard H. Taylor<sup>a</sup>, Shidong Wang<sup>a</sup>, Junkai Xue<sup>a</sup>, Kesong Yang<sup>a</sup>, Ohad Levy<sup>d</sup>, Michael J. Mehl<sup>e</sup>, Harold T. Stokes<sup>c</sup>, Denis O. Demchenko<sup>f</sup>, Dane Morgan<sup>g</sup>

<sup>a</sup> Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708

<sup>b</sup> Department of Physics, Duke University, Durham, NC 27708

<sup>c</sup> Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602

<sup>d</sup> Department of Physics, NRCN, P.O. Box 9001, Beer-Sheva 84190, Israel

<sup>e</sup> U.S. Naval Research Laboratory, 4555 Overlook Avenue, SW, Washington DC 20375

<sup>f</sup> Department of Physics, Virginia Commonwealth University, Richmond, VA 23284

<sup>g</sup> Department of Materials Science and Engineering, 1509 University Avenue, Madison, WI 53706

\* Corresponding author email: stefano@duke.edu; tel: +1 919 660 5506; fax: +1 919 6608963

---

## Abstract

Recent advances in computational materials science present novel opportunities for structure discovery and optimization, including uncovering of unsuspected compounds and metastable structures, electronic structure, surface and nano-particle properties. The practical realization of these opportunities requires systematic generation and classification of the relevant computational data by high-throughput methods. In this paper we present AFLOW (Automatic Flow), a software framework for high-throughput calculation of crystal structure properties of alloys, intermetallics and inorganic compounds. The AFLOW software is available for the scientific community on the website of the materials research consortium, [aflowlib.org](http://aflowlib.org). Its geometric and electronic structure analysis and manipulation tools are additionally available for online operation at the same website. The combination of automatic methods and user online interfaces provide a powerful tool for efficient quantum computational materials discovery and characterization.

**Keywords:** High-throughput, Combinatorial Materials Science, Ab initio, AFLOW

---

## 1. Introduction

A new class of software tools to assess material properties has emerged from two parallel theoretical advancements: quantum-mechanical computations based on density functional theory (DFT), and informatics data mining and evolutionary structure screening strategies. Joined together, these methods make it possible to efficiently screen large sets of material structures with many different combinations of elements, compositions and geometrical configurations. The practical realization of such schemes necessitates a *high-throughput* (HT) approach, which involves setting up and performing many *ab initio* calculations and then organizing and analyzing the results with minimal intervention by the user. The HT concept has already

become an effective and efficient tool for materials discovery [1, 2, 3, 4, 5, 6, 7, 8] and development [9, 10, 11, 12, 13]. Examples of computational materials HT applications include combinatorial discovery of superconductors [1], Pareto-optimal search for alloys and catalysts [14, 15], data-mining of quantum calculations applying principle-component analysis to uncover new compounds [5, 6, 7, 16, 17], Kohn-anomalies search in ternary lithium-borides [18, 19, 20], and multi-optimization techniques used for the study of high-temperature reactions in multicomponent hydrides [21, 22, 23].

In its practical implementation, the HT approach uses some sort of automatic optimization technique to screen a library of candidate compounds and to direct further

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>14 NOV 2011</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-2011 to 00-00-2011</b>	
4. TITLE AND SUBTITLE <b>AFLOW: an automatic framework for high-throughput materials discovery</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Duke University, Department of Mechanical Engineering and Materials Science, Durham, NC, 27708</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>in press, Computational Materials Science (2012).</b>					
14. ABSTRACT <b>Recent advances in computational materials science present novel opportunities for structure discovery and optimization including uncovering of unsuspected compounds and metastable structures, electronic structure, surface and nano-particle properties. The practical realization of these opportunities requires systematic generation and classification of the relevant computational data by high-throughput methods. In this paper we present AFLOW (Automatic Flow), a software framework for high-throughput calculation of crystal structure properties of alloys, intermetallics and inorganic compounds. The AFLOW software is available for the scientific community on the website of the materials research consortium, <a href="http://afwlib.org">afwlib.org</a>. Its geometric and electronic structure analysis and manipulation tools are additionally available for online operation at the same website. The combination of automatic methods and user online interfaces provide a powerful tool for efficient quantum computational materials discovery and characterization.</b>					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>12</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

refinements. The library can be a set of alloy prototypes [24, 7] or a database of compounds such as the Pauling File [25] or the ICSD Database [26, 27]. Rather than calculating one target physical quantity over a large number of structures and compositions, a key philosophy of the HT method is to calculate *a priori* as many different quantities as it is computationally feasible. Properties and property correlations are then extracted *a posteriori*.

This paper describes the HT framework AFLOW which we have been developing over the last decade. The code and manuals describing its many operation options are freely available for download at [aflowlib.org/aflow.html](http://aflowlib.org/aflow.html). Some of its capabilities may also be operated interactively online at a dedicated web page [aflowlib.org/awrapper.html](http://aflowlib.org/awrapper.html).

## 2. Software and Structure Database

The high-throughput framework AFLOW, is an assemblage of software tools comprising over 150,000 lines of C++ code. It is written for UNIX systems (GNU-Linux, Mac OSX) with the GNU suite of compilers. Its main features are fully multi-threaded and parallel. It is designed to run on top of any software for structure energy calculation and is currently optimized for first-principles calculations by Vienna Ab Initio Simulation Package (VASP) [28] (porting to other DFT packages, such as Quantum Espresso [29] is underway). The code offers several options for running the DFT package, on either a single structure or on sets of structures, by searching through subfolders containing `aflow.in` input files and running those that have not been calculated yet.

AFLOW makes it possible to automatically calculate a suite of physical observables over a specified class or a large database of structures with little human intervention to set up the input files, run the calculations, and collate and plot the results. It can also be used to assist in the setup of standardized calculations of individual cases. To maximize the usefulness of the large amount of information produced by the HT approach, the data must be generated and represented in a consistent and robust manner. This is especially true when the goal is, as in many materials science applications, to simultaneously optimize multiple properties. For instance, in catalyst design [30] and superconducting materials development [19, 20], both thermodynamics and electronic structure are essential to the effectiveness of the material. If started as a common Unix daemon through the queue of a computer cluster, AFLOW can generate, run,

correct and converge many calculations per day, with minimum human input.

The backbone of the software is the implementation of ground-state energy calculations to identify stable and meta-stable structures. In addition, it addresses phase diagram construction, electronic band structure, phonon-spectra and vibrational free-energy, surface stability and adsorption of contaminants on surfaces, electron-phonon coupling and superconductivity, local atomic environment expansion [31], and cluster expansion. The code is being developed continuously with new applications integrated into it periodically.

One of the most difficult challenges in HT database generation is responding *automatically* to the failure of a calculation. The most common cause is insufficient hardware resources. Precaution must be taken, for example, to estimate the memory requirement of the tasks and group jobs so that they do not hamper each other if running on the same node. The second most common cause of interruption is due to runtime errors of the *ab initio* calculation itself. These errors include inconsistent reciprocal and  $\mathbf{k}$  lattice meshes, inconsistent atom locations, ill-defined geometries, too few electronic bands to include all the electrons in the system, internal inconsistencies arising in the convergence process, and many others. AFLOW is capable of detecting most of these problems, altering the input accordingly and running again. This automatic restarting of runs is achieved by diagnosing the error message, correcting the appropriate parameters, and restarting the calculation.

As of this writing, the AFLOW database includes ~400 experimental prototypes taken primarily from the Pauling File [25] with some additions from the ICSD [26, 27] and the Navy Crystal Lattice database [32]. The most current list of these experimental prototypes can be viewed by either using the command-line version of AFLOW or the online tool described in Sec. 4. Occasionally, fully relaxed calculations by AFLOW (cell volume and shape and the basis atom coordinates inside the cell) turn up as ground state structures that do not yet have known experimental prototypes [7, 33]. When identified, they are also added to the structure database.

The AFLOW database also includes a few million bcc- and fcc-derived superstructures (those containing up to 20 atoms/cell), and a similar number of hcp-derived superstructures (all those containing up to 24 atoms/cell) enumerated formally using the algorithm of Refs. [34] and [35] (a derivative superstructure is a structure whose lattice vectors are multiples of those of a parent lattice and whose atomic basis vectors correspond to *lattice points* of the parent lattice). In practice only a few hun-

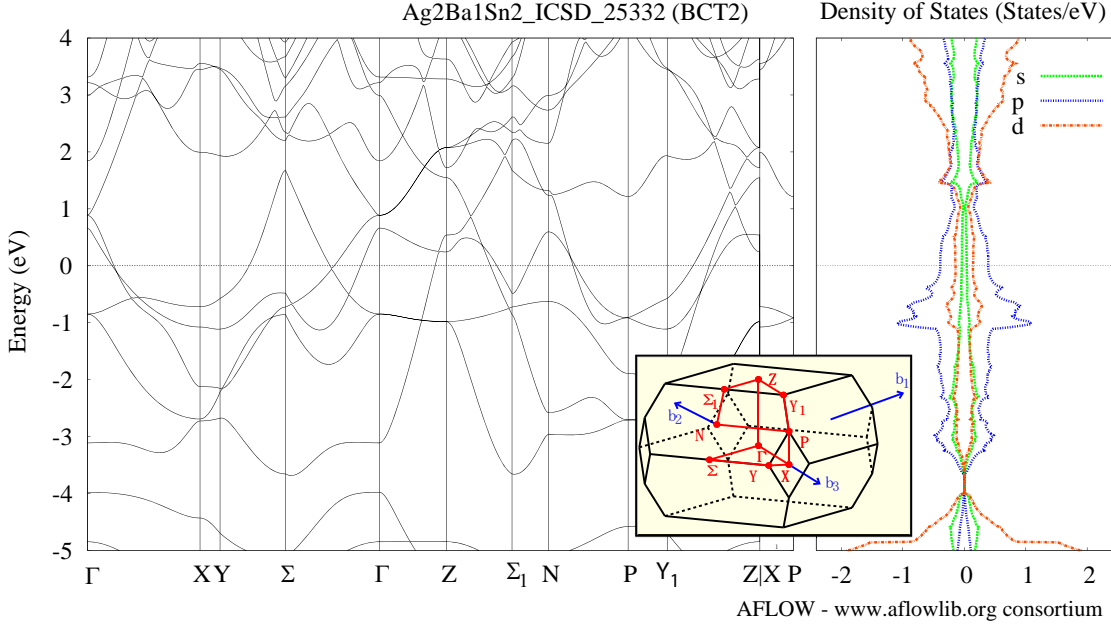


Figure 1: Left: Electronic band structure of  $\text{Ag}_2\text{Ba}_1\text{Sn}_2$ , ICSD structure # 25332, calculated along the Brillouin Zone path shown in the inset. Right: The electronic density of states of the same structure. Both automatically generated using the AFLOW framework.

dred of these structures are usually calculated for each alloy system, but they are useful in constructing cluster expansions [36, 37, 38, 39] which can be employed in synergy with the HT calculations to examine structural configurations more thoroughly than is possible with a purely HT approach [17].

### 3. Basic operation

At the beginning of a calculation, the starting structure, or structures, are selected from the database. The database structural description contains the lattice parameters and atomic positions of the prototype compound. AFLOW then adjusts the lattice parameters (using a weighted average of the pure element volumes) for the specified elements and creates an input file with all the necessary parameters for relaxations, static and band structure runs. This entire procedure requires only the label of a structure in the AFLOW database. It may be performed with equal ease for only one structure with a specific composition or for a large set of structures with many combinations of elements, with a single command. For multiple structures or compositions this command creates a set of subfolders, each including a single input file, through which AFLOW runs automatically until all structures and compositions are calculated.

A HT computational framework must contain a general, reliable, and standardized electronic structure analysis feature. For example, it must automatically determine the Brillouin Zone (BZ) integration path for the 14 Bravais Lattices (with their 24 Brillouin zones) [40] and change the basis (lattice vectors) into a standardized basis so that data can be compared consistently between different projects. Although BZ integration paths have appeared in the literature in the last few decades [41, 42, 43, 44, 45], a standardized definition of the paths for all the different cases has been, to the best of our knowledge, missing, and mistakes in the literature for less-common Brillouin zones are not uncommon. This component of AFLOW was discussed in detail in a separate publication [40].

AFLOW computes structure total energies and electronic band structures using VASP with pseudopotentials and accuracy chosen by the user. As a default, AFLOW employs projector augmented wave (PAW) pseudopotentials with GGA exchange correlation functionals [46, 47] as parameterized by Perdew-Burke-Ernzerhof [48]. Each structure is fully relaxed twice with a convergence tolerance of 1 meV/atom using dense grids of 3,000–6,000  $\mathbf{k}$ -points per reciprocal atom. At the beginning of these structural relaxation steps, a spin-polarized calculation is performed for all structures. If the magnetization is smaller than  $0.025 \mu_B/\text{atom}$ , the spin is turned off for the next relaxations and subsequent calculations to

enhance the calculation speed. After this, the structure is changed into a standard form [40], and another electronic relaxation is performed using fixed coordinates for the lattice vectors and atomic positions. This static run is implemented with a much denser grid of 10,000–30,000 points to get accurate charge densities and density of states for the calculation of the band structure. The Monkhorst-Pack scheme [49] is employed in the grid generation except for hexagonal, fcc and rhombohedral systems in which  $\Gamma$ -centered grid is selected for faster convergence. This process is performed in one calculation, with the single input file AFlow creates as described above. At the completion of such a calculation AFlow invokes appropriate MATLAB or GNUPLOT scripts for data analysis and visualization. All these steps are usually performed automatically, but the user also has the option to operate them through the web interface. An example of a computed electronic band structure and standard BZ path appears in Figure 1.

AFlow’s capability to continuously search subfolders for calculations to run is not limited to DFT calculations. An “alien” mode is also implemented, which allows AFlow to execute other tasks in a high throughput fashion. For instance, the many thousands of grand canonical Monte Carlo calculations used in a recent surface science absorption project [50, 51] were directed and performed by AFlow. In addition, AFlow is equipped with options to run commands or scripts before and after the main program is performed in each folder. This allows AFlow to generate input files on the fly depending on the results of different calculations, so that ad-hoc optimization can be implemented by the user. It also improves the flexibility of recovery from a crash or an unconverged run and increases the overall versatility and throughput of the calculation.

#### 4. Structure analysis tools

AFlow offers structure analysis and manipulation tools which are also useful to users who do not need to perform HT calculations or to create databases for datamining. These users may prepare standard unit cell input files and extract the appropriate  $k$ -points path using the command version of AFlow, called ACONVASP, or the online interface on our website [aflowlib.org/awrapper.html](http://aflowlib.org/awrapper.html), (shown in Figure 2). Runs should then be performed according to the following protocol: Unit cells must first be reduced to standard primitive, then appropriately relaxed by AFlow. Before the static run, the cells should be reduced again to standard primitive, since symmetry and orientation might have changed during the relaxation. The user should

then perform the static run and then project the eigenvalues along the directions which are specified in the “ $k$ -path” option. If the user is running AFlow and VASP, the web interface can also prepare the input file, `afLOW.in`, which will perform all the mentioned tasks. The conversion and analysis operations may be carried on a structure file stored in the database, or supplied by the user in the input box Input POSCAR. The operations implemented on the web interface are the following:

- Normal Primitive; Generates a primitive cell in the most compact form (not necessarily unique).
- Standard Primitive; Chooses a primitive cell such that the Wigner-Seitz cell defined by the reciprocal vectors coincides with one of the possible 24 Brillouin zones [42, 41, 40].
- Standard Conventional; Generates a conventional unit cell (not necessarily primitive).
- Minkowski lattice reduction; Generates a maximally-compact cell (not necessarily unique).
- Niggli Standardized form; Generates a cell that conforms to the Niggli standard form [52].
- WYCKOFF-CAR/ABCCAR to POSCAR; Generates a POSCAR file from a structure file containing the standard crystallographic information.
- POSCAR to ABCCAR; Generates a structure file containing lattice parameters (rather than lattice vectors) from a POSCAR file.
- Bring atoms in the cell; Remaps atomic coordinates to lie inside the unit cell.
- Cartesian coordinates; Converts the atomic coordinates in a structure file to Cartesian coordinates (from lattice coordinates).
- Fractional coordinates; Converts atomic positions from fractional coordinates (i.e, lattice or direct coordinates) to Cartesian coordinates.
- Data and Extended Data; Generates lattice parameters (cell lengths and angles) and other information (volume, reciprocal lattice, symmetry information, etc.).
- Symmetry Information; Lattice Type of the crystal and the lattice, Pearson symbol, Space group and Wyckoff positions, Point Group Lattice Matrices, Point Group Crystal Matrices and factor Group Crystal Matrices and Translations.

☐ aflow/aconvasp VERSION   ☐ aflow HELP   ☐ aconvasp HELP   ☐ apennsy HELP  
☐ Normal Primitive (fast)  
☐ Standard Primitive (slow) [DOI: 10.1016/j.commatsci.2010.05.010]  
☐ Standard Conventional (slow) [DOI: 10.1016/j.commatsci.2010.05.010]  
☐ Minkowski lattice reduction   ☐ Niggli Standardized form  
☐ WYCKOFF-CAR/ABCCAR to POSCAR (read aconvasp.pdf for definition of wyckcar/abccar)  
☐ POSCAR to ABCCAR  
☐ Bring atoms in the cell   ☐ Cartesian coordinates   ☐ Fractional coordinates  
☐ Data (a,b,c,alpha,beta,gamma)   ☐ Extended Data (a,b,c,alpha,beta,gamma lattices, point group)  
☐ Lattice Type of the crystal   ☐ Lattice Type of the lattice   ☐ Pearson symbol  
☐ Space Group (PLATON)   ☒ New Space Group (FINDSYM)   ☐ New Findsym output  
☐ Point Group Lattice Matrices   ☐ Point Group Crystal Matrices  
☐ Factor Group Crystal Matrices and Translations   ☐ Identical atoms   ☐ Sites Symmetries  
☐ Kpath in the reciprocal space for band structure calculations [DOI: 10.1016/j.commatsci.2010.05.010]  
 (with respect to the reciprocal vectors generated from the standard primitive lattice vectors).  
☐ New Interstitial Positions (online up to 8 atoms/cell) [DOI: 10.1103/PhysRevB.79.134203]  
☐ List of prototypes of AFLOW  
☐ List of Bands (reference needed)  
☐ List of Libraries (reference needed)  
☐ Test for development

reference (if required, otherwise leave it empty)

   (be patient)

Input POSCAR (put your poscar):

```

CuPt/309 - (309) - CuPt/ Ca/Ge [AB7] (309) (htgc library)
-122.124500
0.0000000000000000 1.0000000000000000 1.0000000000000000
1.0000000000000000 0.0000000000000000 1.0000000000000000
1.0000000000000000 1.0000000000000000 0.0000000000000000
1 7
Direct (8)
0.0000000000000000 0.0000000000000000 0.0000000000000000 Cu
0.5000000000000000 0.0000000000000000 0.0000000000000000 Pt
0.0000000000000000 0.5000000000000000 0.0000000000000000 Pt
0.0000000000000000 0.0000000000000000 0.5000000000000000 Pt
0.0000000000000000 0.5000000000000000 0.5000000000000000 Pt
0.5000000000000000 0.0000000000000000 0.5000000000000000 Pt
0.5000000000000000 0.5000000000000000 0.0000000000000000 Pt
0.5000000000000000 0.5000000000000000 0.5000000000000000 Pt

```

Figure 2: Input interface for the structure analysis tools of AFLOW. Users may apply these tools to a structure included in the database or enter their own structure in the input box. The tools include symmetry analysis, format conversions, and transforming structures between equivalent representations (primitive versus conventional, etc.).

- Identical atoms and site symmetries; Identifies which atoms in a structure correspond to the same Wyckoff positions.
- K-path in the reciprocal space; Provides directions in  $\mathbf{k}$ -space needed for band structure calculations and provides a picture of the corresponding Brillouin zone and path.
- Interstitial Cages Positions; provides all the geometric locations of interstitials in the structure.

The user interface also provides access to the software manuals, under the buttons marked aflow HELP, aconvasp HELP and apennsy HELP, and to the current list of experimental prototypes in the AFLOW database under List of prototypes of AFLOW.

## 5. Example Applications

### 5.1. Vibration spectra and free energy

AFLOW can calculate phonon dispersion curves using three different approaches: the direct force constant method, following the definition of Ref. [54, 55] and description of Ref. [56], the linear response method

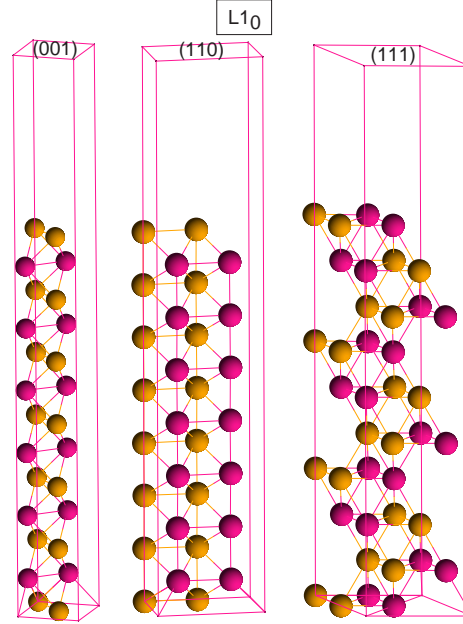


Figure 4: The supercells of (001), (110), and (111) slabs of fcc binary  $L_{10}$  structure built by ACONVASP.

for PAWs [57], and the frozen phonon method as implemented in the FROZSL code of H. Stokes and L. Boyer [58, 59, 60].

In the first method, the atomic displacements necessary to fully determine the dynamical matrix are prescribed automatically, the forces are calculated by vasp and mapped onto symmetry-equivalent directions by AFLOW. The method is general and can treat low-symmetry structures as easily as high-symmetry systems, minimizing user-time overhead. To reduce computational efforts, the symmetry of the supercells is fully employed and only the unique atoms are distorted along the minimum independent directions required for the construction of the force constant matrix. When necessary, we use linear response [57] for the non-analytical part of the dynamical matrix [61, 62], to reproduce the correct LO-TO splitting (splitting between longitudinal and transverse optical phonon frequencies). For the linear response method as implemented in vasp, AFLOW sets up the input files, runs and converges the calculation automatically. AFLOW uses the generated dispersion curves to calculate the vibrational contribution to the free energy, entropy and specific heat (see example in Figure 3). The whole task is performed by the APL library of AFLOW.

In the frozen-phonon approach, AFLOW-FROZSL, the user sets up an aflow.in input file containing instruc-

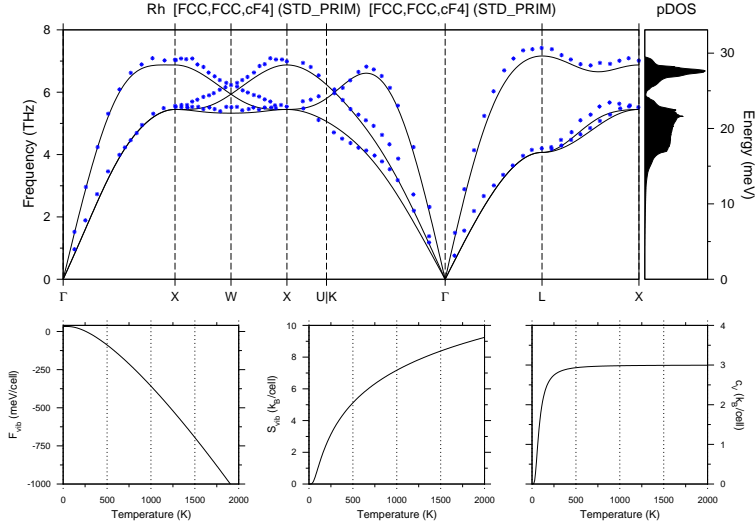


Figure 3: Top: Phonon dispersion curves for face-centered cubic rhodium. Solid lines are computed frequencies, dots are experimental results [53]. The phonon density of states is shown on the right hand side. Bottom: Vibrational free energy, vibrational entropy, and specific heat calculated automatically using the AFLOW framework.

tions in the FROZSL format [60]. To perform the calculation AFLOW repeatedly calls FROZSL to generate all the irreducible representations of the phonons, calculate the various geometries and the phonon spectrum in the requested  $\mathbf{k}$  points or paths. The code is multi-threaded: the various irreducible representations can be tackled simultaneously in a computer cluster while an AFLOW daemon waits for their completion to produce the spectrum. For instance the several thousand calculations of Ref. [63] were rapidly addressed by this method.

We chose to implement all three approaches for convenience. While the first gives more information in the whole Brillouin Zone, the second and the third can be used to rapidly scan spectra at particular  $\mathbf{k}$  points, such as when searching for Kohn anomalies [20]. Currently, the FROZSL/FINDSYM source packages are maintained by the AFLOW team. They are included in the AFLOW distribution and have been modified to compile with the GNU suite.

### 5.2. Design of high-index surfaces in complex multi-component compounds

Surface segregation, adsorption, chemisorption, catalytic reaction and other surface phenomena are of ultimate importance in many areas of modern technology, including catalysis, corrosion protection, batteries, electronics etc. In most theoretical studies of surface effects, it is necessary to know the exact positions of surface and subsurface atoms. In cases of low-index surfaces and simple compounds, the construction of a surface can be

done manually. However, the technologically interesting high-index surfaces [64, 65] and multicomponent compounds with complex underlying crystal lattices require an automated tool.

The input file used for the surface construction describes the initial bulk crystal structure, denoted by the corresponding AFLOW database label. The structure may have an arbitrary complex crystal lattice with any number of atom types. The generated surface file contains the complete data (both direct and Cartesian coordinates) for the constructed surface supercell. The Miller indices ( $hkl$ ) are based on the Bravais lattice that corresponds to the unit cell of the bulk crystal structure. An example of (001), (110), and (111) supercell slabs of the fcc binary  $L1_0$  structure are shown in Figure 4. AFLOW can easily build a supercell containing a predetermined number  $N_f$  of atomic planes with given Miller indices ( $hkl$ ) of a given crystal structure. A supercell may also contain a number  $N_e$  of “empty” ( $hkl$ )-planes, without actual atoms. In the last case, a periodically repeated supercell represents the slabs separated by empty space. Such slab construction allows one to study surface effects, e.g. surface energy, surface stress, and surface segregation, from first principles [66, 67, 68, 69, 70, 65]. The obtained surface data may also be effectively used in simulation of nanoparticles [70, 65]. The variation of  $N_f$  and  $N_e$  is important for studying convergence with respect to slab width and inter slab distance, respectively. Within ( $hkl$ )-



planes, the constructed supercell consists of  $p_1 \times p_2$  two-dimensional primitive unit cells, where  $p_1, p_2 = 1, 2, \dots$  are specified. Large  $p_1$  and  $p_2$  are used to introduce surface defects, e.g. vacancy or solute atoms, far enough from each other to diminish their interaction. For instance, all the surfaces studied in Ref. [65] were produced with this method.

### 5.3. Nanoparticle generation

Starting from an input crystal structure, a specified radius and required separation, ACONVASP generates a structure file of a nano-particle of that radius made of the same lattice and separated as specified from its nearest neighbors. The origin of the particle may be set to the origin of the unit cell, an atom in the crystal structure or any point in Cartesian or fractional coordinates. The radius and separation distance are in Å. This option replaces the cumbersome manual generation of nano-particle structure files, usually employed in studies of such particles, and should enable investigation of large sets of nano-particles in a high-throughput fashion

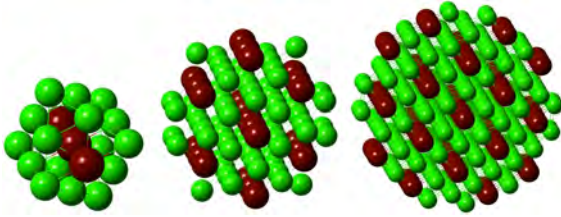


Figure 5: Nano-particles based on the  $L1_3$  structure built by ACONVASP, with radii of 5 Å (left), 8 Å (center) and 12 Å (right).

### 5.4. Topological identification of interstitial sites

An additional feature of AFLOW allows users to identify interstitial sites inside any crystallographic structure. Using as input a file specifying the atomic positions in the structure unit cell, it returns the location, radius and coordination of each site. The algorithm identifies quadruplets of non-coplanar atoms, where the first atom belongs to the unit cell and the others are separated by less than the longest diagonal of the unit cell. *Cages* are defined by spheres touching all four atoms of the quadruplet that do not contain any atoms. An *interstitial* position is found if a center of a cage is inside the unit cell. By considering all the possible combinations, symmetrically inequivalent interstitials can be identified through calculation of their site symmetry (with the space group of the unit cell). Note that in unit cells with complex arrangements, many of the interstitial positions can be extremely close. Thus, adjacent interstitial atoms

located in any of those close positions could deform the nearby local atomic environment and *relax* to the same final location. Hence, the number of symmetrically inequivalent cages can be further reduced by *agglomeration* of such a set of positions into a single interstitial site, upon insertion of interstitial atoms.

Given the host cell geometry and the interstitial species, the occupation of the final irreducible cages can be automatically simulated by AFLOW, which calculates their energies, entropy and solubility in the small concentration limit [71, 72]. These operations are implemented in a multi-threaded manner to accelerate the calculation in a multi-core environment. Figure 7 presents an example of the output of a multi-threaded interstitial position search performed online through our servers ([aflowlib.org/awrapper.html](http://aflowlib.org/awrapper.html)), for the  $AgZr_3-L1_3$  structure, using the interface of Figure 2. Here AFLOW finds 14 cages: 6 octahedral and 8 tetrahedral. Reduction through the space group leads to only 5 irreducible positions, two octahedral positions with  $r = 2.2225$  Å ( $Ag_1Zr_5$  and  $Ag_2Zr_4$ ) and three tetrahedral positions with  $r = 1.9248$  Å ( $Ag_1Zr_3$ ,  $Ag_2Zr_2$  and  $Zr_4$ ). The equivalence and the multiplicity per cell are provided in the text. Figure 6 shows these octahedral and tetrahedral interstitials in the standard conventional representation of this structure  $L1_3$ . For clarity, the conventional oP8 structure is shown, instead of the primitive oS8. The two unique octahedrons are blue and the three unique tetrahedrons are in pink. The interstitial atom is in the center of the polytopes.

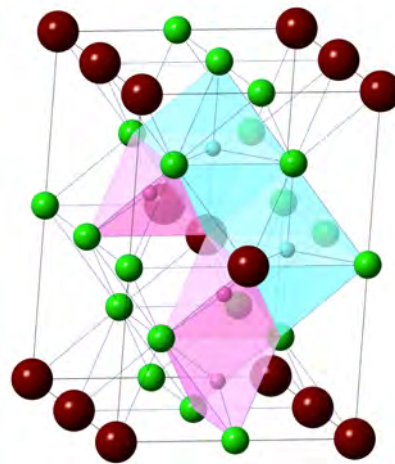


Figure 6: 6-coordinated (blue) and 4-coordinated (pink) interstitial sites in the  $L1_3$  structure, found in a topological search by ACONVASP.



```

-----
afLOW VERSION 30308: [Stefano Curtarolo - 2003-2012]
-----
CAGES CALCULATION - Stefano Curtarolo
-----
str.pgroup.size()=8
str.agroup.size()=4
str.fgroup.size()=8
str.pgroup_xtal.size()=8
str.fgroup.size()=4212
-----
REFERENCE STRUCTURE BELOW
-----
AgZr/11 - (11) - FCC [AB3] (11) (htgc library)
1.000000
3.143228 5.444232 5.444232 109.4712 90.0000 90.00000
Direct(4) [AB3]
0.000000 0.000000 0.000000 Ag 00[00]*_(1)
0.500000 0.000000 0.500000 Zr 01[01]*_(2)
0.500000 0.500000 0.000000 Zr 02[01]
0.000000 0.500000 0.500000 Zr 03[03]*_(1)
-----
oswrite=1
roughness=1.075000000000000
str.scale=1.000000000000000
size of the grid=108
SCANNING STABLE CAGES (4 points)
x(frac) y(frac) z(frac) S radius(A) T C P coordination thread
0.500000 0.250000 0.250000 X 2.2225 - 6 4 [Ag(0),Zr(1),Zr(2),Zr(3),Ag(0),Zr(3)] ithread=0
0.000000 0.375000 0.125000 X 1.9248 - 4 4 [Ag(0),Zr(1),Zr(3),Zr(2)] ithread=1
0.000000 0.125000 0.375000 X 1.9248 - 4 4 [Ag(0),Zr(1),Zr(3),Zr(1)] ithread=0
0.000000 0.250000 0.750000 X 2.2225 - 6 4 [Zr(1),Zr(3),Zr(1),Zr(2),Ag(0),Zr(2)] ithread=19
0.000000 0.750000 0.250000 X 2.2225 - 6 4 [Zr(2),Zr(3),Zr(2),Zr(1),Ag(0),Zr(1)] ithread=19
1.000000 0.250000 0.750000 X 2.2225 - 6 4 [Zr(1),Zr(3),Zr(1),Zr(3),Ag(0),Zr(2)] ithread=15
0.500000 0.375000 0.625000 X 1.9248 - 4 4 [Zr(1),Zr(3),Zr(2),Zr(3)] ithread=1
0.500000 0.125000 0.875000 X 1.9248 - 4 4 [Zr(1),Ag(0),Zr(2),Ag(0)] ithread=6
1.000000 0.750000 0.125000 X 1.9248 - 6 4 [Zr(2),Zr(1),Zr(2),Zr(3),Ag(0),Zr(1)] ithread=15
0.000000 0.625000 0.875000 X 1.9248 - 4 4 [Zr(3),Zr(2),Zr(2),Ag(0)] ithread=22
0.500000 0.375000 0.625000 X 2.2225 - 4 4 [Zr(2),Ag(0),Zr(1),Ag(0)] ithread=13
0.500000 0.625000 0.375000 X 1.9248 - 4 4 [Zr(2),Zr(3),Zr(1),Zr(2)] ithread=0
0.500000 0.750000 0.750000 X 2.2225 - 6 4 [Zr(3),Zr(2),Zr(1),Ag(0),Zr(3),Ag(0)] ithread=13
0.000000 0.875000 0.625000 X 1.9248 4 4 [Zr(3),Zr(1),Zr(1),Ag(0)] ithread=5
UNIQUE=14 - cage4 radius max=2.2225985 - cage4 radius min=1.9248268 -
REDUCIBLE= 14
-----
PERFORMING REDUCTION:
x(frac) y(frac) z(frac) S radius(A) T C P coordination
1.000000 0.250000 0.750000 X 2.2225 1 6 4 [Zr(1),Zr(2),Zr(1),Zr(3),Ag(0),Zr(2)]
0.500000 0.250000 0.750000 X 2.2225 2 6 4 [Ag(0),Zr(1),Zr(2),Zr(3),Ag(0),Zr(3)]
0.000000 0.625000 0.875000 X 1.9248 3 4 4 [Zr(3),Zr(2),Zr(2),Ag(0)]
0.500000 0.375000 0.625000 X 1.9248 4 4 4 [Zr(1),Ag(0),Zr(2),Zr(3)]
0.500000 0.125000 0.875000 X 1.9248 5 4 4 [Zr(1),Ag(0),Zr(2),Ag(0)]
ISREDUCIBLE= 5
-----
ISREDUCIBLE/REDUCIBLE summary: 12/38
x(frac) y(frac) z(frac) S radius(A) T C P coordination
1.000000 0.250000 0.750000 X 2.2225 1 6 4 [Zr(1),Zr(2),Zr(1),Zr(3),Ag(0),Zr(2)]
equiv 1.000000 0.750000 0.250000 type=1
equiv 0.000000 0.250000 0.750000 type=1
equiv 0.000000 0.750000 0.250000 type=1
0.500000 0.250000 0.750000 X 2.2225 2 6 4 [Ag(0),Zr(1),Zr(2),Zr(3),Ag(0),Zr(3)]
equiv 0.500000 0.750000 0.250000 type=2
0.000000 0.625000 0.875000 X 1.9248 3 4 4 [Zr(3),Zr(2),Zr(2),Ag(0)]
equiv 0.000000 0.875000 0.625000 type=3
equiv 0.000000 0.375000 0.625000 X 1.9248 type=3
equiv 0.000000 0.125000 0.875000 type=3
0.500000 0.375000 0.625000 X 1.9248 3 4 4 [Zr(1),Zr(3),Zr(2),Zr(3)]
equiv 0.500000 0.625000 0.375000 type=4
0.500000 0.125000 0.875000 X 1.9248 5 4 4 [Zr(1),Ag(0),Zr(2),Ag(0)]
equiv 0.500000 0.875000 0.125000 type=5
-----
ISREDUCIBLE=12 REDUCIBLE=38
-----
afLOW VERSION 30308: [Stefano Curtarolo - 2003-2010]
-----
[afLOW] *****
[afLOW] CAGES IN AFLOW IN FORMAT
[afLOW] *****
[VASP_POSCAR_MODE_EXPLICIT]START_cages_T1
AgZr/11 - (11) - FCC [AB3] (11) (htgc library) R=2.2226 T=1 C=4 P=4 [Zr(1),Zr(2),Zr(1),Zr(3),Ag(0),Zr(2)]
1.000000
3.143228 5.444232 5.444232 109.4712 90.0000 90.00000
1 1 3
Direct(5) [AB1C3]
0.000000 0.000000 0.000000 Ag
0.000000 0.250000 0.750000 H
0.500000 0.000000 0.500000 Zr
0.500000 0.500000 0.000000 Zr
0.000000 0.500000 0.500000 Zr
[VASP_POSCAR_MODE_EXPLICIT]STOP_cages_T1
[afLOW] *****
[VASP_POSCAR_MODE_EXPLICIT]START_cages_T2
AgZr/11 - (11) - FCC [AB3] (11) (htgc library) R=2.2226 T=2 C=4 P=4 [Ag(0),Zr(1),Zr(2),Zr(3),Ag(0),Zr(3)]
1.000000
3.143228 5.444232 5.444232 109.4712 90.0000 90.00000
1 1 3
Direct(5) [AB1C3]
0.000000 0.000000 0.000000 Ag
0.000000 0.250000 0.750000 H
0.500000 0.000000 0.500000 Zr
0.500000 0.500000 0.000000 Zr
0.000000 0.500000 0.500000 Zr
[VASP_POSCAR_MODE_EXPLICIT]STOP_cages_T2
[afLOW] *****
[VASP_POSCAR_MODE_EXPLICIT]START_cages_T3
AgZr/11 - (11) - FCC [AB3] (11) (htgc library) R=1.92483 T=3 C=4 P=4 [Zr(3),Zr(2),Zr(2),Ag(0)]
1.000000
3.143228 5.444232 5.444232 109.4712 90.0000 90.00000
1 1 3
Direct(5) [AB1C3]
0.000000 0.000000 0.000000 Ag
0.000000 0.625000 0.875000 H
0.500000 0.000000 0.500000 Zr
0.500000 0.500000 0.000000 Zr
0.000000 0.500000 0.500000 Zr
[VASP_POSCAR_MODE_EXPLICIT]STOP_cages_T3
[afLOW] *****
[VASP_POSCAR_MODE_EXPLICIT]START_cages_T4
AgZr/11 - (11) - FCC [AB3] (11) (htgc library) R=1.92483 T=4 C=4 P=4 [Zr(1),Zr(3),Zr(2),Zr(3)]
1.000000
3.143228 5.444232 5.444232 109.4712 90.0000 90.00000
1 1 3
Direct(5) [AB1C3]
0.000000 0.000000 0.000000 Ag
0.000000 0.375000 0.625000 H
0.500000 0.000000 0.500000 Zr
0.500000 0.500000 0.000000 Zr
0.000000 0.500000 0.500000 Zr
[VASP_POSCAR_MODE_EXPLICIT]STOP_cages_T4
[afLOW] *****
[VASP_POSCAR_MODE_EXPLICIT]START_cages_T5
AgZr/11 - (11) - FCC [AB3] (11) (htgc library) R=1.92483 T=5 C=4 P=4 [Zr(1),Ag(0),Zr(2),Ag(0)]
1.000000
3.143228 5.444232 5.444232 109.4712 90.0000 90.00000
1 1 3
Direct(5) [AB1C3]
0.000000 0.000000 0.000000 Ag
0.000000 0.125000 0.875000 H
0.500000 0.000000 0.500000 Zr
0.500000 0.500000 0.000000 Zr
0.000000 0.500000 0.500000 Zr
[VASP_POSCAR_MODE_EXPLICIT]STOP_cages_T5
[afLOW] *****

```

Figure 7: Output of a multi-threaded search of interstitial cages in the L1<sub>3</sub> structure, showing two octahedral positions with  $r = 2.2225\text{\AA}$  (Ag<sub>1</sub>Zr<sub>5</sub> and Ag<sub>2</sub>Zr<sub>4</sub>) and three tetrahedral positions with  $r = 1.9248\text{\AA}$  (Ag<sub>1</sub>Zr<sub>3</sub>, Ag<sub>2</sub>Zr<sub>2</sub> and Zr<sub>4</sub>). The cell geometries including these interstitials are given in the bottom part of the printout.

### 5.5. APENNSY: automatic analysis

The results of high-throughput AFLOW runs are analyzed and manipulated for further processing by the

APENNSY module. The basic operation of this module is the production of a system-specific data file that includes information on all the computed structures, their energies, the input and output (after relaxation) crystal prototype and space group, and the binary convex hull. It requires as input the output files generated by the AFLOW run for the computed system.

For example, in a binary alloy system, the *convex hull* at zero-temperature is used to create the phase diagram. To construct the convex hull, APENNSY finds the minimum energy structure at each composition, and calculates their formation enthalpies

$$H_f = E(x) - xE_A - (1 - x)E_B, \quad (1)$$

where  $E_A$  and  $E_B$  are the energies of the pure elements and  $E(x)$  is the energy of the alloy with concentration  $x$  of element A. APENNSY then identifies the extremal structures, i.e., those that lie below the tie-lines connecting their neighbors at adjacent compositions. These extremal structures are the stable structures in the binary system and constitute its convex hull. APENNSY plots the convex hull automatically using GNUPLOT or MATLAB.

The module can also be used to print the computed data in formats appropriate for use as input for other codes. For example, it can print all the bcc (or fcc or hcp)-based structures with the total energy per unit cell or energy per atom. This is useful for cluster expansion codes such as UNCLE [39]. Another format is used for automatic miscibility determination. All these data files can be generated off-line by operating APENNSY on a library of input files, or on-line by selecting a specific system and the requested option at the AFLOW website [afLOWlib.org](http://afLOWlib.org) (“binary alloy library” entry). Additional options of APENNSY provide data that is external to the structure calculations, such as predictions of system miscibility following Miedema and Hume-Rothery mixing rules [73, 74].

## 6. Future developments: High-throughput hybrid functionals calculations

The local (LDA) and semi-local (GGA) approximations to the DFT, expanded with “+U” techniques [75], have been incredibly successful in describing a wide variety of materials properties, particularly those related to the ground state. In some cases, however, the drawbacks of these approximations present an insurmountable obstacle. One of the most illustrative areas is the energetics of defects in wide band-gap semiconductors. For example, in the case of oxygen vacancies in ZnO, the defect formation energies and thermodynamic

transition levels obtained by different LDA/GGA approaches vary by several eV [76]. Similar problems exist in DFT calculations of optical properties of insulators and semiconductors, properties of correlated materials, reaction energetics and adsorption of small molecules [77, 78]. An affordable and accurate approach to overcome these problems does not exist yet. The state of the art many-body methods, such as self-consistent GW [79] exhibit excellent agreement with experiments, but suffer from high computational cost. The hybrid functional methods[80] represent a pragmatic compromise between the local and nonlocal many-body approaches. The results of hybrid functional calculations are often in excellent agreement with computationally expensive cutting edge many-body methods, yet favorable scaling allows routine calculations of a few hundred atoms in the unit cell [81].

In a hybrid functional calculation the LDA exchange correlation part of the density functional is mixed with a Fock-type exchange part in varying proportions. For example, the popular Perdew-Burke-Ernzerhof hybrid functional (PBE0) contains 25% of the exact exchange, 75% of GGA exchange, and 100% of GGA correlation energy [82]. In a related Heyd-Scuseria-Ernzerhof (HSE06) functional the Fock exchange interactions are separated into long- and short-range parts. The short range part includes 25% of exact exchange and 75% of semi-local GGA exchange, while the long-ranged part is replaced with an approximate semi-local GGA expression [83]. The splitting is accomplished by introducing the screening of the exchange interactions (similar to the screened exchange approach) with optimal screening length of approximately 7-10 Å. This approach presents a middle case between the GGA calculation (no exact exchange) and the PBE0 hybrid functional (all exchange is long ranged). The HSE06 functional eliminates some unphysical features of the exact exchange approach. It also has the computational advantage of better convergence of the long-ranged exchange part, since screening out the long range exact exchange greatly reduces the cost of calculating the non-local exchange interaction. One also often finds it useful to tune the amount of exact exchange for a particular material (the so called  $\alpha$ -tuned hybrid functionals) in order to obtain the best agreement of computed band-gap with experiment. This is particularly useful in cases where the correct value of the band-gap is critical. For example, in the calculations of defect energetics, tuning the band-gap of the host material in the bulk and then running calculations for supercells containing impurities and defects has been shown to produce results in very good agreement with experiment [84].

The hybrid functional calculations are computationally more demanding than LDA/GGA calculations, due to the need to fully calculate the non-local Fock exchange integrals. For example, a bulk ZnO HSE06 calculation ( $8 \times 8 \times 6$   $\Gamma$ -centered  $\mathbf{k}$ -point mesh, with 40 irreducible  $\mathbf{k}$ -points, 400 eV cutoff) using VASP takes approximately 5 hours on 8 CPU's. The same set up for the GGA/PBE (+U) calculation takes less than a minute. The memory requirements in this case, for a HSE06 calculation, are also increased by a factor of five compared to those of GGA. The scaling of the computational cost varies for different hybrid functionals, and is approximately  $O(N^{2.5})$  for small system sizes, and linear for HSE06 beyond 15 Å and for PBE0 beyond 100 Å. Using plane waves the scaling of hybrid functional methods is  $(N_{\text{bands}}N_k)^2N_{\text{FFT}}\log(N_{\text{FFT}})$ , which is approximately linear with the number of atoms in the bulk [81]. The standard (semi-) local DFT(+U) methods using plane waves normally scale as  $O(N^3)$ .

### 6.1. Framework of DFT+U and Hybrid Functional calculations

**HT-DFT+U:** Within the high-throughput framework, AFLOW currently employs the DFT+U technique and allows the user to choose the appropriate parameters. As default, AFLOW takes advantage of the Dudarev [85] formalism within GGA+U. The values of  $U$  are listed in Table 1.

Table 1: Values of  $U_{\text{eff}}$  parameters in eV for the Dudarev GGA+U approach implemented in AFLOW. From Refs. [86, 40] and references therein.

Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga
4.4	2.7	3.5	4.0	4.6	5.0	5.1	4.0	7.5	3.9
Nb	Mo	Tc	Ru	Rh	Pd			Cd	In
2.1	2.4	2.7	3.0	3.3	3.6			2.1	1.9
Ta	W	Re	Os	Ir	Pt				
2.0	2.2	2.4	2.6	2.8	3.0				
La	Ce	Pr	Gd	Nd	Sm	Eu	Tm	Yb	Lu
7.5	6.3	5.5	6.0	6.2	6.4	5.4	6.0	6.3	3.8

**HT-DFT+Hybrid:** We are currently expanding the AFLOW framework to perform HSE06 calculations, and we present some preliminary results. The extension is not trivial: the calculations of electronic structure using hybrid functionals differ from the standard LDA/GGA ones, with or without “+U”. One cannot compute the bands in the familiar non-selfconsistent way of the LDA/GGA because non-local exchange is not determined by the pre-computed charge density. Therefore to obtain the eigenvalues for strings of  $\mathbf{k}$ -points the following recipe is being implemented. First, AFLOW performs

a standard LDA/GGA calculation. Second, a hybrid functional run is performed starting from the LDA/GGA wavefunctions on the same  $\mathbf{k}$ -point mesh and energy cutoff chosen by AFLOW. The number of bands is dynamically adjusted to achieve full convergence. Then, to facilitate efficient use of computational resources, it is further adjusted to include only a few bands over the highest occupied state. Third, the electronic structure is computed by performing a hybrid functional run explicitly defining the same  $\mathbf{k}$ -point mesh in addition to the desired  $\mathbf{k}$ -path as specified in the high symmetry examples of Ref. [40]. The crucial step here is to set the mixing weights of the extra  $\mathbf{k}$ -path to zero, while keeping the original mesh intact. Since the orbitals at the extra  $\mathbf{k}$ -path do not contribute to the total energy, and the wavefunctions on the original mesh are converged as input, it is only necessary to converge the orbitals at the extra  $\mathbf{k}$ -points mesh with the appropriate VASP instructions. The band structure calculation following this recipe will take approximately the same amount of time as a regular total energy hybrid calculation since the orbitals at the standard mesh are pre-converged.

As an example Figure 8 shows the HSE06 band structure computed along the standard high symmetry lines. Comparisons can be drawn with the PBE+U electronic structure of Ref. [87]. The HSE06 calculation used the standard value of  $\alpha = 0.25$ . (Note that by using  $\alpha = 0.375$  the band-gap can be brought to agreement with experiment.) The value of the band-gap is considerably improved  $E_g = 2.48$  eV in comparison with GGA and PBE+U values of  $E_g = 0.7$  eV and  $E_g = 1.82$  eV, respectively (the phenomenologically corrected value is 3.36 eV [86]). In other materials this improvement is often better, since ZnO produces one of the largest LDA/GGA band-gap errors. The HSE06 band-gaps for narrow and medium gap semiconductors, with the standard ratio of the Fock exchange, are in good agreement with experiment. For wide band-gap semiconductors and insulators the gaps, although underestimated, present a significant improvement over LDA/GGA values [80]. In addition, hybrid functionals improve effective mass estimation for almost all systems, typically yielding values within a few percent of experiment [88].

A fully functional, consistent and robust AFLOW framework with hybrid functionals is planned for 2012.

## 7. Summary

We describe the AFLOW software package for HT calculations of material properties. It should be helpful to the materials scientist seeking to determine properties

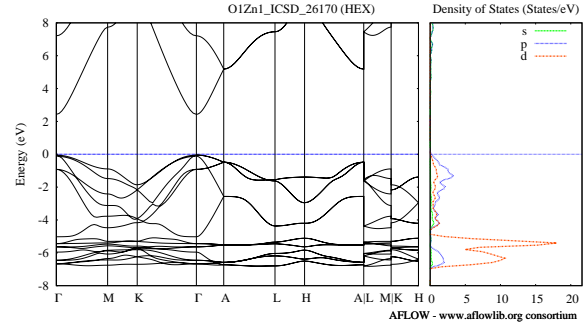


Figure 8: HSE06 hybrid functional band structure of ZnO. The HSE06 gap is 2.48 eV. LDA/GGA gap is 0.7 eV while the PBE+U value is 1.82 eV [87]. Standard HSE06 parameters were used, the ratio of exact exchange is 0.25 and the screening parameter is  $0.2 \text{ \AA}^{-1}$ .

of alloy and compound structures. The code, and operation manuals describing its features, are freely available for download at [aflowlib.org/aflow.html](http://aflowlib.org/aflow.html). The structure manipulation and analysis capabilities of AFLOW are also available online at the ACONVASP-online entry of this website.

## 8. Acknowledgments

The authors acknowledge Gerbrand Ceder, Marco Buongiorno-Nardelli, Leeor Kronik, Natalio Mingo, and Stefano Sanvito for fruitful discussions. Research supported by ONR (N00014-11-1-0136, N00014-10-1-0436, N00014-09-1-0921), NSF (DMR-0639822, DMR-0908753), and the Department of Homeland Security - Domestic Nuclear Detection Office. We are grateful for extensive use of the Fulton Supercomputer Center at Brigham Young University and Teragrid resources (MCA-07S005). SC acknowledges support by the Feinberg fellowship at the Weizmann Institute of Science.

## References

- [1] X. D. Xiang, X. Sun, G. Briceno, Y. Lou, K.-A. Wang, H. Chang, W. G. Wallace-Freedman, S.-W. Chen, P. G. Schultz, *Science* 268 (1995) 1738–1740.
- [2] Y.-M. Chiang, D. R. Sadoway, M. K. Aydinol, Y.-I. Jang, B. Huang, G. Ceder, *Nature* 392 (1998) 694.
- [3] G. H. Jóhannesson, T. Bligaard, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, J. K. Nørskov, *Phys. Rev. Lett.* 88 (2002) 255506.
- [4] D. P. Stucke, V. H. Crespi, *Nano Lett.* 3 (2003) 1183.
- [5] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, G. Ceder, *Phys. Rev. Lett.* 91 (2003) 135503.
- [6] D. Morgan, G. Ceder, S. Curtarolo, *Meas. Sci. Technol.* 16 (2005) 296.
- [7] S. Curtarolo, D. Morgan, G. Ceder, *Calphad* 29 (2005) 163.

- [8] C. C. Fischer, K. J. Tibbetts, D. Morgan, G. Ceder, *Nat. Mater.* 5 (2006) 641.
- [9] H. Koinuma, I. Takeuchi, *Nat. Mater.* 3 (2004) 429–438.
- [10] J. L. Spivack, J. N. Cawse, D. W. Whisenhunt, B. F. Johnson, K. V. Shalyaev, J. Male, E. J. Pressman, J. Y. Ofori, G. L. Soloveichik, B. P. Patel, T. L. Chuck, D. J. Smith, T. M. Jordan, M. R. Brennan, R. J. Kilmer, E. D. Williams, *Applied Catalysis A: General* 254 (2003) 5 – 25.
- [11] T. R. Boussie, G. M. Diamond, C. Goh, K. A. Hall, A. M. La-Pointe, M. Leclerc, C. Lund, V. Murphy, J. A. W. Shoemaker, U. Tracht, H. Turner, J. Zhang, T. Uno, R. K. Rosén, J. C. Stevens, *J. Am. Chem. Soc.* 125 (2003) 4306–4317.
- [12] R. A. Potyrailo, B. J. Chisholm, W. G. Morris, J. N. Cawse, W. P. Flanagan, L. Hassib, C. A. Molaison, K. Ezbiasky, G. Medford, H. Reitz, *J. Comput. Chem.* 5 (2003) 472–478.
- [13] R. A. Potyrailo, I. Takeuchi, *Measurement Science and Technology* 16 (2005) 1.
- [14] T. Bligaard, G. H. Johansson, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, J. K. Nørskov, *Appl. Phys. Lett.* 83 (2003) 4527.
- [15] M. P. Anderson, T. Blidgaard, K. E. L. A. Kustov, J. Greeley, T. Johansson, C. H. Christensen, J. K. Nørskov, *J. Catal.* 239 (2006) 501.
- [16] O. Levy, R. V. Chepulsii, G. L. W. Hart, S. Curtarolo, *J. Am. Chem. Soc.* 132 (2010) 833.
- [17] O. Levy, G. L. W. Hart, S. Curtarolo, *J. Am. Chem. Soc.* 132 (2010) 4830.
- [18] A. N. Kolmogorov, S. Curtarolo, *Phys. Rev. B* 73 (2006) 180501(R).
- [19] M. Calandra, A. N. Kolmogorov, S. Curtarolo, *Phys. Rev. B* 75 (2007) 144506.
- [20] A. N. Kolmogorov, M. Calandra, S. Curtarolo, *Phys. Rev. B* 78 (2008) 094520.
- [21] C. Wolverton, D. J. Siegel, A. R. Akbarzadeh, V. Ozolins, *J. Phys.: Condens. Matt.* 20 (2008) 064228.
- [22] D. J. Siegel, C. Wolverton, V. Ozolins, *Phys. Rev. B* 76 (2007) 134102.
- [23] A. R. Akbarzadeh, V. Ozolins, C. Wolverton, *Advanced Materials* 19 (2007) 3233–3239.
- [24] T. B. Massalski, H. Okamoto, P. R. Subramanian, L. Kacprzak (Eds.), *Binary Alloy Phase Diagrams*, American Society for Metals, Materials Park, OH, 1990.
- [25] P. Villars, M. Berndt, K. Brandenburg, K. Cenual, J. Daams, F. Hulliger, T. Massalski, H. Okamoto, K. Osaki, A. Prince, H. Putz, S. Iwata, *J. Alloys Compound.* 367 (2004) 293–297.
- [26] A. D. Mighell, V. L. Karen, *Acta Cryst.* a49 (1993) c409.
- [27] A. Belsky, M. Hellenbrandt, V. L. Karen, P. Luksch, *Acta Cryst.* b58 (2002) 364–369.
- [28] G. Kresse, J. Hafner, *Phys. Rev. B* 47 (1993) 558–561.
- [29] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, R. M. Wentzcovitch, *J. Phys.: Condens. Matt.* 21 (2009) 395502.
- [30] J. K. Nørskov, T. Bligaard, J. Rossmeisl, C. H. Christensen, *Nat. Chem.* 1 (2009) 37.
- [31] P. Villars, in: J. H. Westbrook, R. L. Fleisher (Eds.), *Crystal Structures of Intermetallic Compounds*, Wiley, New York, 1994, pp. 1–49.
- [32] *Naval Research Laboratory Crystal Structure database*, <http://cst-www.nrl.navy.mil/lattice/> (2011).
- [33] O. Levy, G. L. W. Hart, S. Curtarolo, *Phys. Rev. B* 81 (2010) 174106.
- [34] G. L. W. Hart, R. W. Forcade, *Phys. Rev. B* 77 (2008) 224115.
- [35] G. L. W. Hart, R. W. Forcade, *Phys. Rev. B* 80 (2009) 014120.
- [36] J. M. Sanchez, F. Ducastelle, D. Gratias, *Physica A* 128 (1984) 334–350.
- [37] D. de Fontaine, *Cluster Approach to Order-disorder Transformations in Alloys*, volume 47 of *Solid State Physics*, academic Press, New York, 1994.
- [38] G. L. W. Hart, V. Blum, M. J. Walorski, A. Zunger, *Nat. Mater.* 4 (2005) 391.
- [39] D. Lerch, O. Wieckhorst, G. L. W. Hart, R. W. Forcade, S. Müller, *Model. Simul. Mater. Sci. Eng.* 17 (2009) 055003.
- [40] W. Setyawan, S. Curtarolo, *Comp. Mat. Sci.* 49 (2010) 299.
- [41] C. J. Bradley, A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids: Representation Theory for Point Groups and Space Groups*, Clarendon Press, Oxford, 1972.
- [42] G. Burns, A. M. Glazer, *Space Groups for Solid State Scientists*, Academic Press, Boston, 1990.
- [43] S. C. Miller, W. F. Love, *Tables of Irreducible Representations of Space Groups and Co-representations of Magnetic Groups*, Pruett Press, Boulder, 1967.
- [44] O. V. Kovalev, *Irreducible Representations of the Space Groups*, Gordon and Breach, New York, 1965.
- [45] A. Casher, M. Glucky, Y. Gur, *The Irreducible Representations of Space Groups*, W. A. Benjamin Inc., New York, 1969.
- [46] W. Kohn, L. J. Sham, *Phys. Rev.* 140 (1965) A1133.
- [47] P. E. Blöchl, *Phys. Rev. B* 50 (1994) 17953.
- [48] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 77 (1996) 3865.
- [49] H. J. Monkhorst, J. D. Pack, *Phys. Rev. B* 13 (1976) 5188.
- [50] W. Setyawan, R. D. Diehl, S. Curtarolo, *Phys. Rev. Lett.* 102 (2009) 055501.
- [51] R. D. Diehl, W. Setyawan, S. Curtarolo, *J. Phys.: Condens. Matt.* 20 (2008) 314007.
- [52] P. Niggli, *Handbuch Der Experimentalphysik*, volume 7(1), Akademische Verlagsgesellschaft, 1928.
- [53] A. Eichler, K.-P. Bohnen, W. Reichardt, J. Hafner, *Phys. Rev. B* 57 (1998) 324.
- [54] A. A. Maradudin, E. W. Montroll, G. H. Weiss, I. P. Ipatova, *Theory of Lattice Dynamics in the Harmonic Approximation*, Academic Press, New York, 1971.
- [55] O. Madelung, *Introduction to Solid-state Theory*, Springer-Verlag, 3 edition, 1996.
- [56] G. Kresse, J. Furthmüller, J. Hafner, *J. Phys.: Condens. Matt.* 9 (1997) 7861.
- [57] M. Gajdoš, K. Hummer, G. Kresse, J. Furthmüller, F. Bechstedt, *Phys. Rev. B* 73 (2006) 045112.
- [58] H. T. Stokes, *Ferroelectrics* 164 (1995) 183–188.
- [59] L. L. Boyer, H. T. S. M. J., Mehl, *Ferroelectrics* 164 (1995) 177–181.
- [60] H. T. Stokes, L. L. Boyer, *FROZSL*, [www.physics.byu.edu/~stokesh/isotropy.html](http://www.physics.byu.edu/~stokesh/isotropy.html) (2002).
- [61] X. Gonze, C. Lee, *Phys. Rev. B* 55 (1997) 10355–10368.
- [62] Y. Wang, J. J. Wang, W. Y. Wang, Z. G. Mei, S. L. Shang, L. Q. Chen, Z. K. Liu, *J. Phys.: Condens. Matter* 22 (2010) 202201.
- [63] K. M. Poduska, L. Regev, E. Boaretto, L. Addadi, S. Weiner, L. Kronik, S. Curtarolo, *Advanced Materials* 23 (2011) 550–554.
- [64] N. Tian, Z.-Y. Zhou, S.-G. Sun, Y. Ding, Z. L. Wang, *Science* 316 (2007) 732–735.
- [65] R. V. Chepulsii, S. Curtarolo, *ACS Nano* 5 (2011) 247–254.
- [66] V. Fiorentini, M. Methfessel, *J. Phys.: Condens. Matt.* 8 (1996) 6525–6529.
- [67] A. V. Ruban, H. L. Skriver, J. K. Nørskov, *Phys. Rev. B* 59

- (1999) 15990–16000.
- [68] A. U. Nilekar, A. V. Ruban, M. Mavrikakis, *Surf. Sci.* 603 (2009) 91 – 96.
  - [69] N. E. Singh-Miller, N. Marzari, *Phys. Rev. B* 80 (2009) 235407.
  - [70] R. V. Chepulsii, W. H. Butler, A. van de Walle, S. Curtarolo, *Scr. Mater.* 62 (2010) 179–182.
  - [71] R. V. Chepulsii, S. Curtarolo, *Phys. Rev. B* 79 (2009) 134203.
  - [72] R. V. Chepulsii, S. Curtarolo, *Acta Materialia* 57 (2009) 5314.
  - [73] A. R. Miedema, P. F. de Chatel, F. R. de Boer, *Physica B&C* 100B (1980) 1.
  - [74] W. Hume-Rothery, *The Metallic State*, Oxford University Press, Oxford, 1931.
  - [75] V. I. Anisimov, F. Aryasetiawan, A. I. Lichtenstein, *J. Phys.: Condens. Matt.* 9 (1997) 767.
  - [76] S. Lany, A. Zunger, *Phys. Rev. B* 78 (2008) 235104.
  - [77] L. A. Curtiss, K. Raghavachari, P. C. Redfern, J. A. Pople, *J. Chem. Phys.* 106 (1997) 1063.
  - [78] G. Kresse, A. Gil, P. Sautet, *Phys. Rev. B* 68 (2003) 073401.
  - [79] M. Shishkin, M. Marsman, G. Kresse, *Phys. Rev. Lett.* 99 (2007) 246403.
  - [80] M. Marsman, J. Paier, A. Stroppa, G. Kresse, *J. Phys.: Condens. Matt.* 20 (2008) 064201.
  - [81] J. Paier, R. Hirschl, M. Marsman, G. Kresse, *J. Chem. Phys.* 122 (2005) 234102.
  - [82] J. P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* 105 (1996) 9982.
  - [83] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* 118 (2003) 8207.
  - [84] F. Oba, A. Togo, I. Tanaka, J. Paier, G. Kresse, *Phys. Rev. B* 77 (2008) 245202.
  - [85] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, A. P. Sutton, *Phys. Rev. B* 57 (1998) 1505.
  - [86] W. Setyawan, R. M. Gaumé, S. Lam, R. S. Feigelson, S. Curtarolo, *ACS Comb. Sci.* 13 (2011) 382–390.
  - [87] [www.aflowlib.org/material.php?proto\\_name=O1Zn1\\_ICSD.26170](http://www.aflowlib.org/material.php?proto_name=O1Zn1_ICSD.26170)  
ZnO ICSD #26170.
  - [88] Y.-S. Kim, M. Marsman, G. Kresse, F. Tran, P. Blaha, *Phys. Rev. B* 82 (2010) 205212.